16.4 Adaptive Frequency and Biasing Techniques for Tolerance to Dynamic Temperature-Voltage Variations and Aging

James Tschanz[1], Nam Sung Kim[1], Saurabh Dighe[1], Jason Howard[1], Gregory Ruhl[1], Sriram Vangal[1], Siva Narendra[2], Yatin Hoskote[1], Howard Wilson[1], Carol Lam[1], Matthew Shuman[4], Carlos Tokunaga[3], Dinesh Somasekhar[1], Stephen Tang[1], David Finan[1], Tanay Karnik[1], Nitin Borkar[1], Nasser Kurd[1], Vivek De[1]

[1]Intel, Hillsboro, OR, [2]Tyfone, Portland, OR
[3]University of Michigan, Ann Arbor, MI, [4]Oregon State University, Corvallis OR

The maximum operating frequency (F_{max}) of a processor is traditionally set at a constant value based on the maximum operating temperature (T_{max}) and supply voltage (V_{cc}) droops. In addition, the expected slowdown of the operating speed during processor lifetime due to transistor aging is used as a guardband for setting the F_{max}. The maximum operating voltage (V_{max}) is constrained by transistor reliability at T_{max}, as well as cooling and current delivery requirements. However, during normal operating conditions, the hot-spot temperature is not always at the maximum, and V_{cc} droops are not always present. Voltage-temperature variations during normal operation and the degree of aging also vary from processor to processor, depending on usage history and process variations. Therefore, designs with fixed frequency and V_{cc} are suboptimal under typical usage conditions. Dynamic voltage scaling [1-3], sometimes used in conjunction with dynamic body bias [4-5], improves efficiency by allowing a processor to adapt its voltage, frequency, and body bias to a changing workload. The recently reported Foxton technology [6-7] adjusts frequency and V_{cc} in response to power and temperature variations, exploiting the power headrooms available during runtime to boost average performance. In this paper, we explore schemes to dynamically adapt various combinations of frequency, V_{cc} and body bias to: 1) changes in temperature, 2) supply noises, and 3) transistor aging, to maximize average performance or improve energy efficiency.

A testchip containing a TCP offload accelerator core [8], a data input buffer, V_{cc} droop sensors [9], thermal sensors, a dynamic adaptive biasing (DAB) controller, distributed noise injectors, body bias generators, and a 3-PLL dynamic frequency unit, is implemented in 90nm CMOS (Figs. 16.4.1 & 16.4.7). The DAB controller receives inputs from the thermal sensors and droop detectors. Average supply current is sensed by the off-chip voltage regulator module (VRM), and digitally communicated to the DAB controller on chip. The programmable noise injectors generate various supply noises and load currents, in addition to those generated by the core during normal operation. The DAB controller drives the dynamic frequency unit, body bias generators, and voltage setting of the off-chip VRM to dynamically adapt frequency, body bias and V_{cc} to achieve optimum settings for the given conditions. The DAB controller is based on a lookup table (Fig. 16.4.2), which is indexed by the output of the thermal, droop, and current sensors and is loaded with pre-characterized data representing the optimum V_{cc}, body bias, and frequency for each of the sensor combinations. The control also includes programmable timers and logic to ensure that transitions in V_{cc}, body bias, and frequency happen in the correct sequence needed for fault-free operation, and to eliminate instability around the sensor trip points. The control is designed to be fast enough to respond to 2nd and 3rd droops in voltage as well as changes in temperature and overall chip activity factor.

Responding to the relatively fast V_{cc} droops also requires a method for changing frequency quickly without waiting for a PLL to re-lock. The clocking subsystem, shown in Fig. 16.4.3, contains 3 PLLs running at independent frequencies and a multiplexer to select between them in a single cycle while ensuring that there are no shortened clock cycles. Several algorithms (Fig. 16.4.3) for changing frequency by switching between multiple PLLs are implemented as part of the frequency control, including a simple algorithm in (a) which switches between 3 locked PLLs, to a flexible algorithm in (c) which keeps one PLL always locked at a frequency higher and lower than the current frequency. When a frequency change is requested, a switch is made to the slower (or faster) PLL, and then the other two PLLs are re-locked and the process repeated. This allows the entire frequency space to be covered in 3% steps. The dynamic frequency algorithms are implemented in the DAB control, and commands are sent to the PLL block to switch between PLLs and update PLL divider values. Clock gating is also implemented to reduce active power consumption of the core when the TCP/IP header has finished processing and the core is idle. Both NMOS and PMOS body bias generators are implemented on the die and each includes a central bias generator (CBG) which is controlled by the DAB control, and many local bias generators (LBGs) distributed throughout the die (Fig. 16.4.2). The PMOS bias implementation includes a differential difference amplifier (DDA), which enables generation of both reverse and forward bias with 32mV resolution. The NMOS bias implementation uses a simple matched source-follower LBG for forward body bias only. Input header data to the core is supplied from the on-chip input buffer, and all arrays and programmable features are loaded through JTAG scan.

The maximum frequency of the design ranges from 2.2GHz at 1V to 3.4GHz at 1.4V and total power consumption at 1.2V is 1.3W for a high-activity test. Frequency is increased by 9 to 22% through application of NMOS and PMOS forward body bias. F_{max} and power are measured across a range of voltages, body biases, and temperatures and the results loaded into the DAB control lookup table. The dynamic response of the chip to temperature changes during a high-workload test (Fig. 16.4.4) shows that while the worst-case frequency is set by the highest expected temperature, as the temperature drops, the core frequency can be increased. At the same time, at low temperature the leakage component of power is reduced, and forward body bias (in this example, NMOS forward body bias) can be applied to further increase the performance. This combination reduces the guardband needed for maximum temperature and, in this example, results in a 1.4% increase in average frequency over the duration of the test. In a similar way, dynamic response to voltage droops (Fig. 16.4.5) allows the droop guardband to be reduced or removed, resulting in increased average performance or reduced power consumption—in this instance, increasing the average frequency by 5%.

Dynamic frequency and body bias capabilities also allow the design to respond to frequency degradation resulting from device aging mechanisms such as NBTI [10]. The threshold voltage increase in the PMOS devices due to aging are compensated by applying increasing amounts of PMOS forward body bias over the lifetime of the part. The measurements of Fig. 16.4.6 show that this technique allows the frequency to remain constant as the part ages, significantly reducing the frequency guardband to account for aging.

Figure 16.4.7 shows the testchip, which is fabricated on a 15mm² die in 90nm dual-V_T CMOS, and is packaged in a 478-pin FCBGA package. The DAB control and additional two PLLs required for dynamic adaptation account for a 5% die area overhead for this technique.

Acknowledgement:
We thank the prototype design team and mask designers of circuits research lab for physical design and chip implementation support; ATD for chip assembly and packaging; Matthew Haycock and Shekhar Borkar for encouragement and support for this work.

References:
[1] T. Burd, T.A. Pering, A.J. Stratakos, and R.W. Brodersen, "A Dynamic Voltage Scaled Microprocessor System," *ISSCC Dig. Tech. Papers*, Feb., 2000.
[2] K.J. Nowka, G.D. Carpenter, E.W. MacDonald, et. al., "A 32-bit PowerPC System-on-a-Chip with Support for Dynamic Voltage Scaling and Dynamic Frequency Scaling," *IEEE J. Solid-State Circuits*, pp. 1441-1447, vol. 37, no. 11, Nov., 2002.
[3] M. Nakai, S. Akui, K. Seno, et. al., "Dynamic Voltage and Frequency Management for a Low-Power Embedded Processor," *IEEE. J. Solid-State Circuits*, pp. 28-35, vol. 40, no. 1, Jan., 2005.
[4] T. Kuroda, K. Suzuki, S. Mita, et. al. "Variable Supply-Voltage Scheme for Low-Power High-Speed CMOS Digital Design," *IEEE J. Solid-State Circuits*, pp. 454-462, vol. 33, no. 3, Mar., 1998.
[5] J. Kao, M. Miyazaki, and A.R. Chandrakasan, "A 175-mV Multiply-Accumulate Unit Using An Adaptive Supply Voltage and Body Bias Architecture," *IEEE J. Solid-State Circuits*, pp. 1545-1554 vol. 37, no. 11, Nov., 2002.
[6] T. Fischer, J. Desai, B. Doyle, et al, "A 90-nm Variable Frequency Clock System for A Power-Managed Itanium Architecture Processor," *IEEE J. Solid-State Circuits,* pp. 218-228, vol. 41, no. 1, Jan., 2006.
[7] R. McGowen, C.A. Poirier, C. Bostak, et al., "Power and Temperature Control on A 90-nm Itanium Family Processor," *IEEE J. Solid-State Circuits*, pp. 229-237, vol. 41, no. 1, Jan., 2006.
[8] Y. Hoskote, B.A. Bloechel, G.E. Dermer, et al., "A TCP Offload Accelerator for 10 Gb/s Ethernet in 90-nm CMOS," *IEEE J. Solid-State Circuits*, pp. 1866-1874, vol. 38, no. 11, Nov., 2003.
[9] A. Muhtaroglu, G. Taylor and T. Rahal-Arabi, "On-Die Droop Detector for Analog Sensing of Power Supply Noise," *Symp. VLSI Circuits*, pp. 193-196, Jun., 2003.
[10] D.K. Schroder and J.A. Babcock, "Negative Bias Temperature Instability: Road to Cross In Deep Submicron Silicon Semiconductor Manufacturing," *J. Applied Physics*, pp. 1-18, vol. 94, no. 1, Jul., 2003.

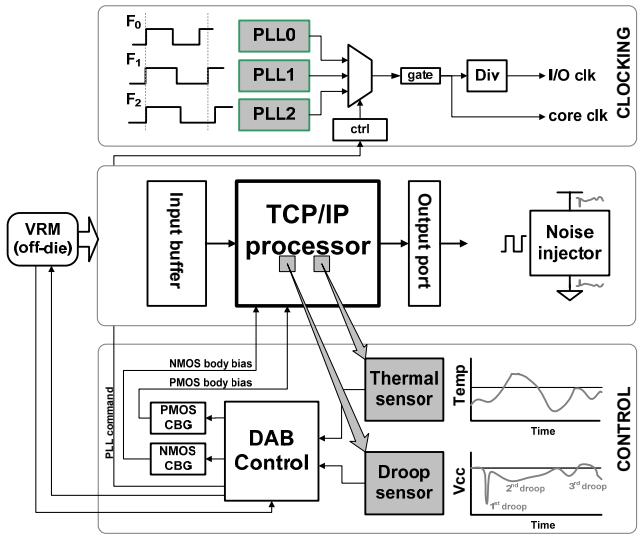


Figure 16.4.1: Block diagram of the testchip, including the TCP/IP processor core, the sensors and dynamic adaptive bias controller, and the dynamic frequency clocking system.

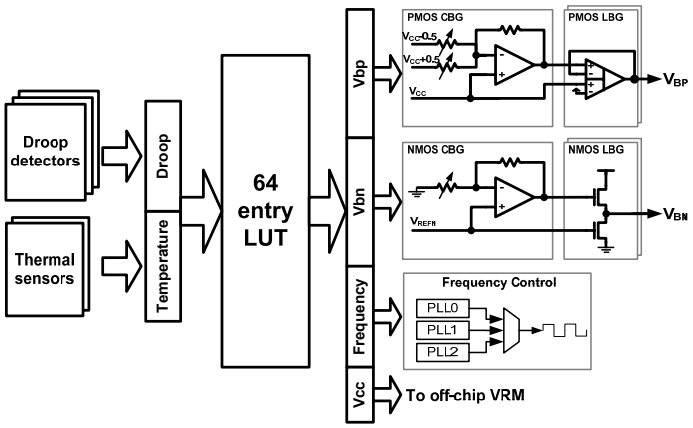


Figure 16.4.2: Operation of the DAB controller. Thermal and droop sensor outputs form the address into a lookup table of pre-characterized frequency, body bias, and supply voltage settings.

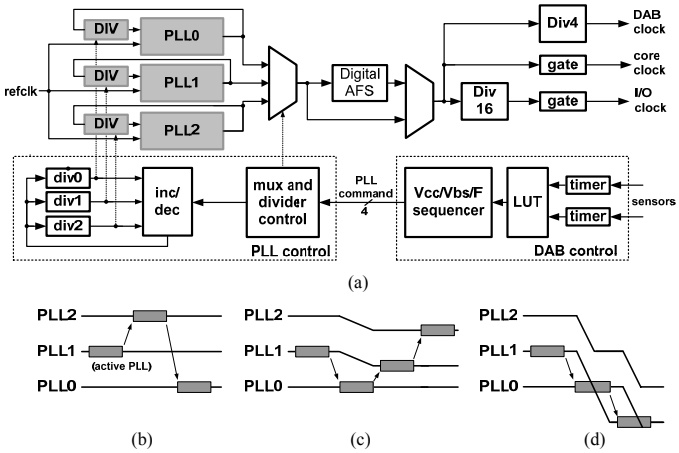


Figure 16.4.3: (a) Components of dynamic frequency system. (b-d) Example frequency change algorithms. (b)Switch between 3 constant-frequency PLLs. (c) Switch to low-frequency PLL and re-lock other two. (d)Re-lock all PLLs one step at a time.

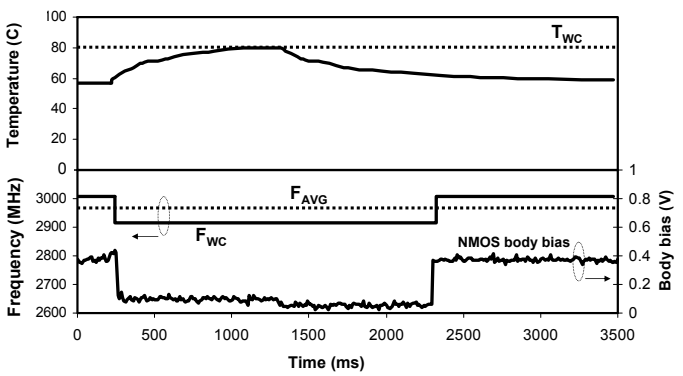


Figure 16.4.4: Adaptive test showing dynamic response of frequency and body bias to a rise in core temperature. Average frequency F_{AVG} is better than the worst-case frequency F_{WC} that is dictated by a fixed-frequency design.

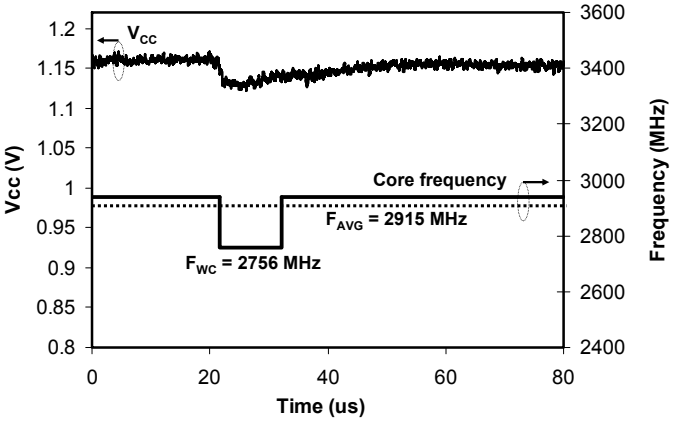


Figure 16.4.5: Adaptive test showing dynamic response of frequency to a supply droop.

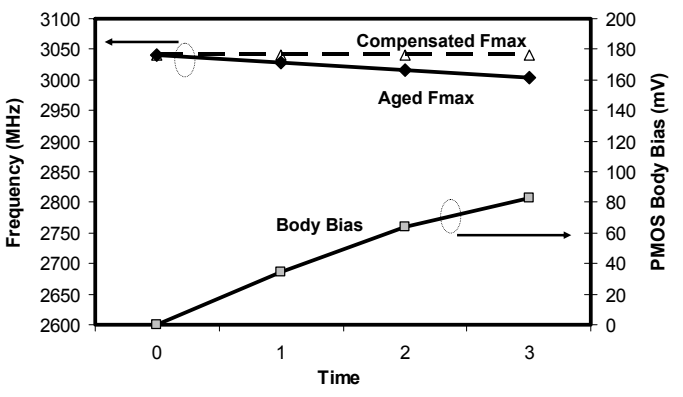
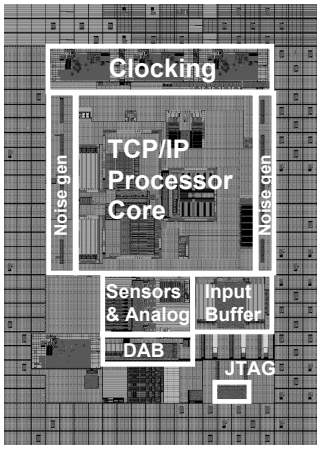


Figure 16.4.6: Compensation of frequency degradation due to aging using PMOS body bias. Dynamic bias allows a constant frequency over the lifetime of the part.

Continued on Page 604

16



Technology	90nm dual-V_T CMOS
Die size	3.3×4.6mm^2
Number of transistors	964K
Fmax	2.9GHz @ 1.2V
Power consumption	1.3W @ 1.2V
On-die body bias	PMOS: RBB & FBB NMOS: FBB
Package	478-pin FCBGA

Figure 16.4.7: Die micrograph and chip characteristics.

